Final Project: Engineering Data Analysis

STAT 4160: Experimental Design

University of Virginia

Adam Chow, Rhiannon Staley

Fall 2023

Problem

Semiconductor manufacturing processes have long and complex assembly flows, so matrix marks and automated 2d-matrix readers are used at several process steps throughout factories. Unreadable matrix marks negatively affect factory run rates because manual entry of part data is required before manufacturing can resume.

A 2^4 factorial experiment was conducted to develop a 2d-matrix laser mark on a metal cover that protects a substrate-mounted die. The design factors are:

1. Laser Power (9 and 13 W)
2. Laser Pulse Frequency (4000 and 12,000 Hz)
3. Matrix Cell Size (0.07 and 0.12 in.)
4. Writing Speed (10 and 20 in./sec.)

The response variable is the unused error correction (UEC). This is a measure of the unused portion of the redundant information embedded in the 2d-matrix. A UEC of 0 represents the lowest reading that still results in a decodable matrix, while a value of 1 is the highest reading. A DMX Verifier was used to measure UEC.

Question

Which factor(s) and/or interactions seem to have the greatest effect on UEC and warrant further study?

Fit the Appropriate Model

2^4 factorial design:

$$\mathbf{Y}_{\mathbf{ijklm}} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\alpha\delta)_{il} + (\beta\gamma)_{jk} + (\beta\delta)_{jl} + (\gamma\delta)_{kl} + (\alpha\beta\gamma)_{ijk} +$$
$$(\alpha\beta\delta)_{ijl} + (\alpha\gamma\delta)_{ikl} + (\beta\gamma\delta)_{jkl} + (\alpha\beta\gamma\delta)_{ijkl} + \epsilon_{ijklm}$$
$$\text{where } \epsilon_{ijklm} \sim N(0, \sigma^2)$$

ANOVA Reduction

To test the significance of our model effects, we will move forward with a half-normal plot which will allow us to estimate the significance of main effects and interactions.

The GLM Procedure

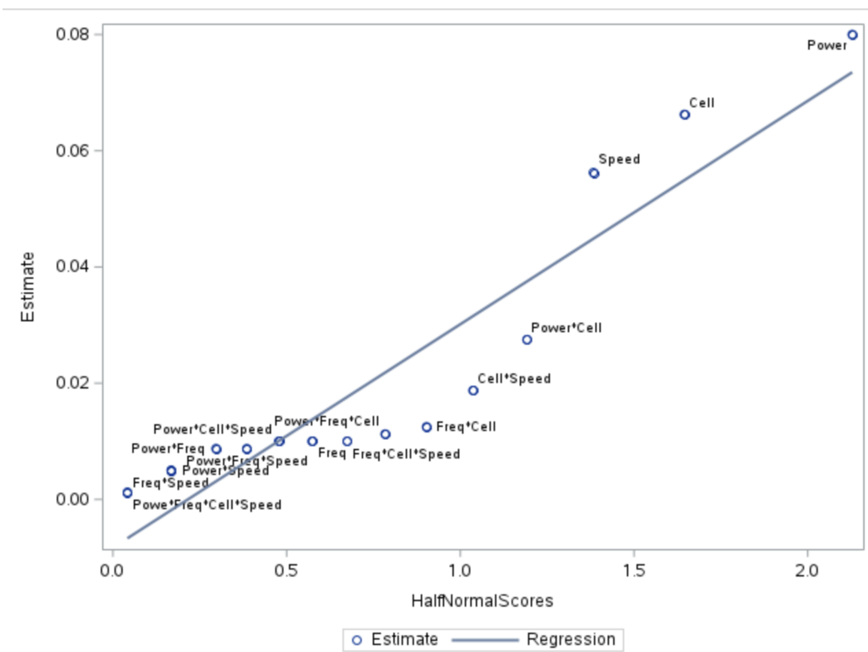
Dependent Variable: UEC

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	0.25357500	0.01690500	.	.
Error	0	0.00000000	.		
Corrected Total	15	0.25357500			

R-Square	Coeff Var	Root MSE	UEC Mean
1.000000	.	.	0.716250

Source	DF	Type I SS	Mean Square	F Value	Pr > F
LaserPower	1	0.10240000	0.10240000	.	.
LaserPulseFrequency	1	0.00160000	0.00160000	.	.
LaserPowe*LaserPulse	1	0.00122500	0.00122500	.	.
MatrixCellSize	1	0.07022500	0.07022500	.	.
LaserPowe*MatrixCell	1	0.01210000	0.01210000	.	.
LaserPuls*MatrixCell	1	0.00250000	0.00250000	.	.
LaserP*LaserP*Matrix	1	0.00202500	0.00202500	.	.
WritingSpeed	1	0.05062500	0.05062500	.	.
LaserPowe*WritingSpe	1	0.00040000	0.00040000	.	.
LaserPuls*WritingSpe	1	0.00040000	0.00040000	.	.
LaserP*LaserP*Writin	1	0.00122500	0.00122500	.	.
MatrixCel*WritingSpe	1	0.00562500	0.00562500	.	.
LaserP*Matrix*Writin	1	0.00160000	0.00160000	.	.
LaserP*Matrix*Writin	1	0.00160000	0.00160000	.	.
Lase*Lase*Matr*Writi	1	0.00002500	0.00002500	.	.

As seen in the output of our ANOVA table there are no F or p-values, because of the lack of replication in this experiment. Because of this, there are no degrees of freedom available to estimate the residual variability (SSE), meaning that we cannot test the significance of any of our model effects.



Looking at the half-normal plot, the effects that appear at first to be significant are LaserPower, MatrixCellSize, WritingSpeed, LaserPower*MatrixCellSize, MatrixCellSize*WritingSpeed, LaserFreq*MatrixCellSize, LaserFreq*MatrixCellSize*WritingSpeed, LaserPower*LaserFreq*MatrixCellSize, and LaserPower*LaserFreq*MatrixCellSize*WritingSpeed.

After plotting the Half-Normal Plot and identifying potentially significant interactions, we have decided to first remove the 4-way interaction to see how much the model changes. We specifically chose the 4-way interaction between LaserPower*LaserFreq*MatrixCellSize*WritingSpeed due to the Effect Hierarchy Principle, which states that lower order interactions are more likely to be important than higher order interactions. When combined with the evidence of insignificance in the half normal plot, this interaction was an obvious removal for our model.

The GLM Procedure

Dependent Variable: UEC

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	14	0.25355000	0.01811071	724.43	0.0291
Error	1	0.00002500	0.00002500		
Corrected Total	15	0.25357500			

R-Square	Coeff Var	Root MSE	UEC Mean
0.999901	0.698080	0.005000	0.716250

Source	DF	Type I SS	Mean Square	F Value	Pr > F
LaserPower	1	0.10240000	0.10240000	4096.00	0.0099
LaserPulseFrequency	1	0.00160000	0.00160000	64.00	0.0792
LaserPowe*LaserPulse	1	0.00122500	0.00122500	49.00	0.0903
MatrixCellSize	1	0.07022500	0.07022500	2809.00	0.0120
LaserPowe*MatrixCell	1	0.01210000	0.01210000	484.00	0.0289
LaserPuls*MatrixCell	1	0.00250000	0.00250000	100.00	0.0635
LaserP*LaserP*Matrix	1	0.00202500	0.00202500	81.00	0.0704
WritingSpeed	1	0.05062500	0.05062500	2025.00	0.0141
LaserPowe*WritingSpe	1	0.00040000	0.00040000	16.00	0.1560
LaserPuls*WritingSpe	1	0.00040000	0.00040000	16.00	0.1560
LaserP*LaserP*Writin	1	0.00122500	0.00122500	49.00	0.0903
MatrixCel*WritingSpe	1	0.00562500	0.00562500	225.00	0.0424
LaserP*Matrix*Writin	1	0.00160000	0.00160000	64.00	0.0792
LaserP*Matrix*Writin	1	0.00160000	0.00160000	64.00	0.0792

After removing the highest order interaction, the Global F-Statistic is 724.23 with a p-value of 0.0291. However, there are still insignificant terms based on p-value like LaserFreq, LaserPower*LaserFreq, LaserFreq*MatrixCellSize, LaserPower*LaserFreq*MatrixCellSize, LaserPower*LaserSped, LaserFreq*WritingSpeed, LaserPower*LaserFreq*WritingSpeed, LaserPower*MatrixCellSize*WritingSpeed, and LaserFreq*MatrixCellSize*WritingSpeed.

At the $\alpha = 0.05$ significance level, all insignificant terms except one, (LaserPower*WritingSpeed) seem to include LaserFrequency, we will remove that from the model and see how the numbers change. We are able to remove all Laser Frequency 3-way interactions, 2-way interaction, and the main effect because the Effect Heredity Principle does not apply in this case, due to the insignificance of the LaserFrequency main effect.

The GLM Procedure

Dependent Variable: UEC

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	0.24297500	0.03471071	26.20	<.0001
Error	8	0.01060000	0.00132500		
Corrected Total	15	0.25357500			

R-Square	Coeff Var	Root MSE	UEC Mean
0.958198	5.082101	0.036401	0.716250

Source	DF	Type I SS	Mean Square	F Value	Pr > F
LaserPower	1	0.10240000	0.10240000	77.28	<.0001
WritingSpeed	1	0.05062500	0.05062500	38.21	0.0003
LaserPowe*WritingSpe	1	0.00040000	0.00040000	0.30	0.5977
MatrixCellSize	1	0.07022500	0.07022500	53.00	<.0001
LaserPowe*MatrixCell	1	0.01210000	0.01210000	9.13	0.0165
WritingSp*MatrixCell	1	0.00562500	0.00562500	4.25	0.0733
LaserP*Writin*Matrix	1	0.00160000	0.00160000	1.21	0.3038

While the Global F-Statistic has significantly decreased, the p-value has also decreased as well. However, there are still a few terms that are insignificant like LaserPower*WritingSpeed, MatrixCellSize*WritingSpeed, and LaserPower*MatrixCellSize*WritingSpeed.

At the $\alpha = 0.05$ significance level, both the 3-way interaction and the 2-way interactions involving WritingSpeed are deemed insignificant, meaning they can all also be removed from the final model.

The GLM Procedure

Dependent Variable: UEC

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	0.23535000	0.05883750	35.51	<.0001
Error	11	0.01822500	0.00165682		
Corrected Total	15	0.25357500			

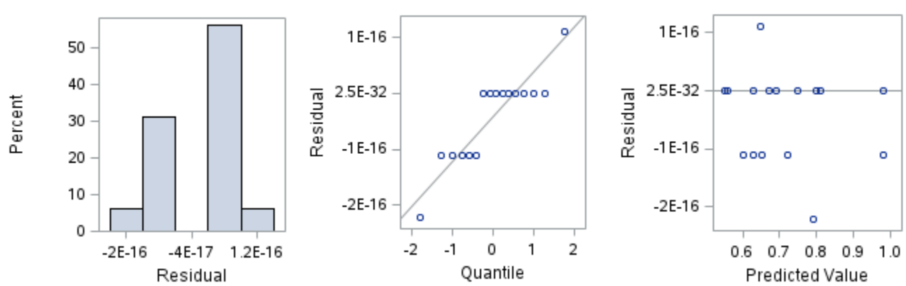
R-Square	Coeff Var	Root MSE	UEC Mean
0.928128	5.682936	0.040704	0.716250

Source	DF	Type I SS	Mean Square	F Value	Pr > F
LaserPower	1	0.10240000	0.10240000	61.81	<.0001
MatrixCellSize	1	0.07022500	0.07022500	42.39	<.0001
LaserPowe*MatrixCell	1	0.01210000	0.01210000	7.30	0.0206
WritingSpeed	1	0.05062500	0.05062500	30.56	0.0002

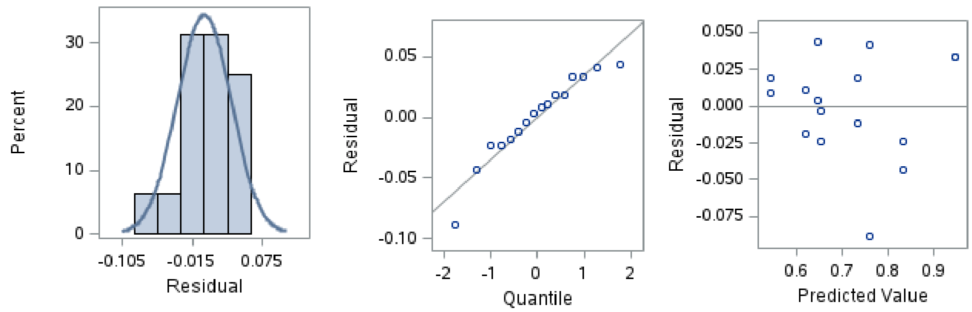
As seen in the output above, this is the best model we've had so far, and is the one we will continue with after checking for assumptions. While we have removed many interaction terms and effects, all of the interactions are now significant with p-values below the α = 0.05 significance level, with the Global F-statistic and p-value following the same trend. Moving forward, we will use this model to finish our analysis.

Verify Assumptions for Inference

Before any reduction to our ANOVA model, this is what the assumptions looked like.



Based on the non-linearity of the QQ-Plot and the less than normal distribution of the histogram, we can assume that this data does not meet the normality assumption. Based on the Residual vs. Predicted plot, the data points do not necessarily indicate any sort of fanning out pattern. Because there is no evidence of us not meeting this assumption, we can assume that this data does meet the assumption of constant variance. After we reduced the ANOVA model to only include LaserPower, CellSize, LaserPower*MatrixCellSize, and WritingSpeed this is what the assumptions look like.



Based on the generally normal distribution of the histogram, and the linearity of the QQPlot, we can assume that this reduced model meets the normality assumption. Based on the Residual vs. Predicted Values plot, we can see that there is no significant fanning out pattern. This means we can assume that the data meets the constant variance assumption. Because our reduced model meets the normality and constant variance assumptions, we do not need to transform the response variable or re-run the ANOVA. Moving forward, we will continue to use the final reduced model.

Pairwise Testing

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey

LaserPower	MatrixCellSize	UEC LSMEAN	LSMEAN Number
-1	-1	0.67500000	1
-1	1	0.59750000	2
1	-1	0.89000000	3
1	1	0.70250000	4

Least Squares Means for Effect LaserPowe*MatrixCell
t for H0: LSMean(i)=LSMean(j) / Pr > |t|

Dependent Variable: UEC

i/j	1	2	3	4
1		2.692646 0.0843	-7.46992 <.0001	-0.95546 0.7765
2	-2.69265 0.0843		-10.1626 <.0001	-3.6481 0.0172
3	7.469921 <.0001	10.16257 <.0001		6.514466 0.0002
4	0.955455 0.7765	3.648101 0.0172	-6.51447 0.0002	

LSMEAN number 3 is the only group number with every pairwise interaction with a p-value less than 0.0002 and the absolute value of the test statistic t being greater than 6.5. This pairwise p-value displays extreme significance: the most significant pair of LaserPower*MatrixCellSize is where LaserPower is 13 Watts (1) and MatrixCellSize is 0.07 inches (-1).

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey

LaserPower	UEC LSMEAN	H0:LSMean1=LSMean2	
		t Value	Pr > |t|
-1	0.63625000	-7.86	<.0001
1	0.79625000		

In the pairwise test on the main effect of the variable LaserPower, it is clear that this main effect is statistically significant with a p-value of <0.0001 and a t value of -7.86. With the negative t value, this indicates that the 9 Watt LaserPower level (-1) is more significant than the 13 Watt level (1).

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey

MatrixCellSize	UEC LSMEAN	H0:LSMean1=LSMean2	
		t Value	Pr > |t|
-1	0.78250000	6.51	<.0001
1	0.65000000		

In the pairwise test on the main effect of the variable MatrixCellSize, it is clear that this main effect is statistically significant with a p-value of <0.0001 and a t value of 6.51. With the positive t value, this indicates that the 0.12 inches MatrixCellSize level (1) is more significant than the 0.07 inches level (-1).

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey

WritingSpeed	UEC LSMEAN	H0:LSMean1=LSMean2	
		t Value	Pr > |t|
-1	0.77250000	5.53	0.0002
1	0.66000000		

In the pairwise test on the main effect of the variable WritingSpeed, it is clear that this main effect is statistically significant with a p-value of 0.0002 and a t value of 5.53. With the positive t value, this indicates that the 20 in./sec WritingSpeed level (1) is more significant than the 10 in./sec level (-1).

Conclusion

Based on the analysis that we have conducted, we have very strong evidence to suggest that the main effects LaserPower, MatrixCellSize, and WritingSpeed and the interaction term LaserPower*MatrixCellSize all have effects on UEC and warrant further study.

Upon further analysis, we determined that LaserPower at the 9 Watt level, MatrixCellSize at 0.12 inches, and WritingSpeed at 20 in./sec are more significant factors in terms of predicting effects on the Unused Error Correction (UEC) of 2-d matrix readers for semiconductor manufacturer processes. This means that these levels of factors have greater effects than their respective counterparts, indicating that these levels of the variables warrant further study.

The specific interaction term MatrixCellSize*LaserPower is most significant at the 0.07 in./13 Watt level, indicating that this level of the interaction warrants further study.